National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Regressions in AIRS v5 Retrieval

Evan Manning

Sung-Yung Lee

California Institute of Technology
Jet Propulsion Laboratory

October 11, 2007

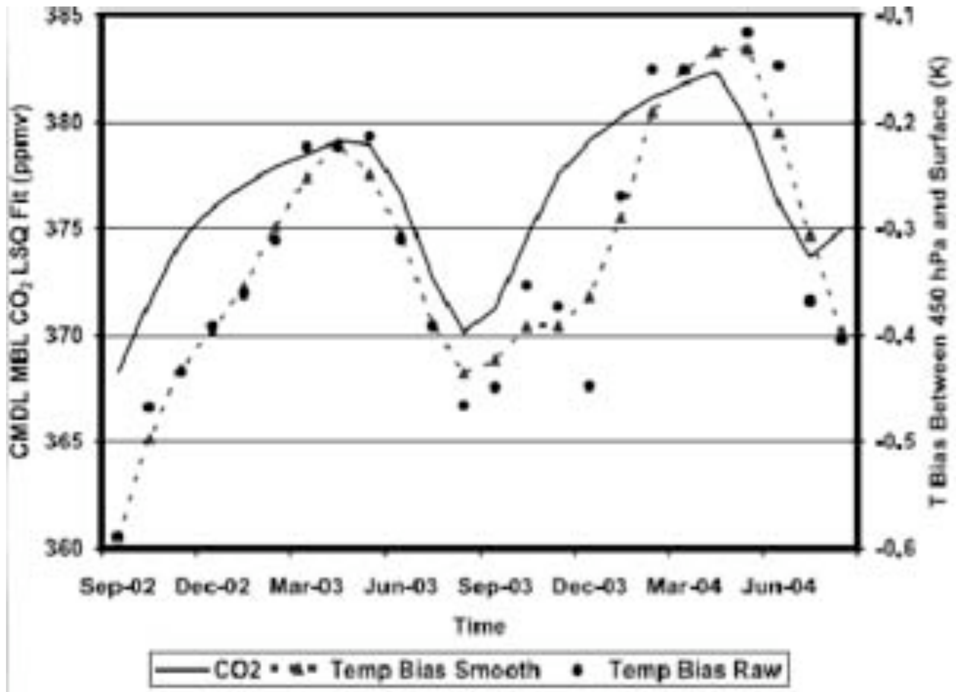
This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration.

Summary

- **Murty Divakarla of NOAA and Thomas Hearty of NASA have shown spurious trends ~100 mK/yr in version 4 & 5 AIRS retrievals vs. truth**
- **Evidence points to regression retrieval steps as a major source of these**
- **Version 6 AIRS retrievals will reduce reliance on regressions and improve practices where regressions are retained**

From Divakarla -- Apparent Trend in AIRS v4 vs. Radiosonde

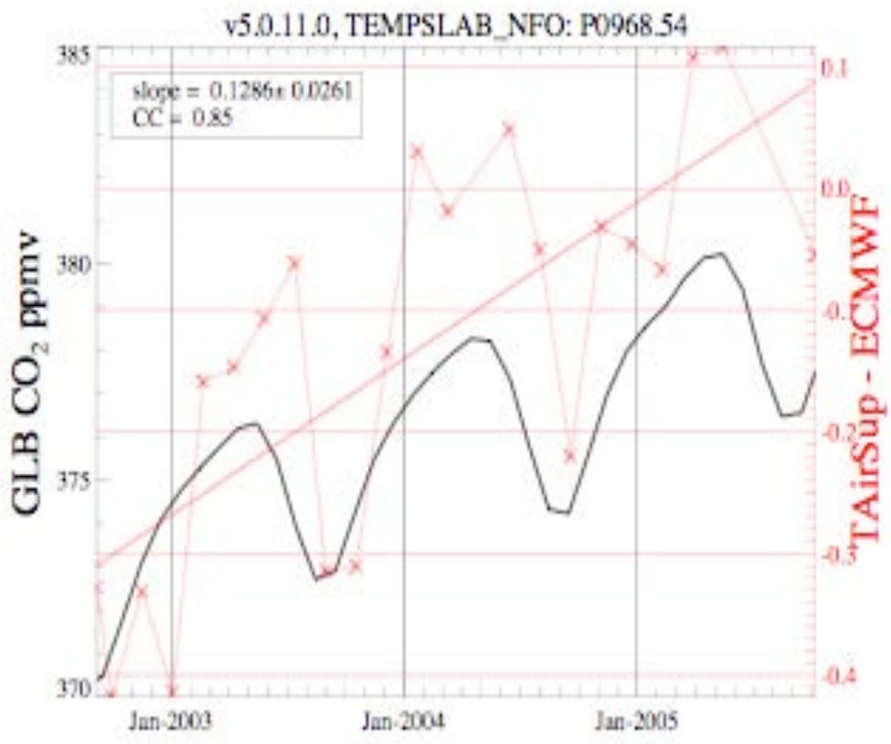
- **Divakarla et al 2006**
- **Correlated with CO_2**
- **AIRS version 4**
- **AIRS version 5 added changing CO_2 background in physical retrieval**

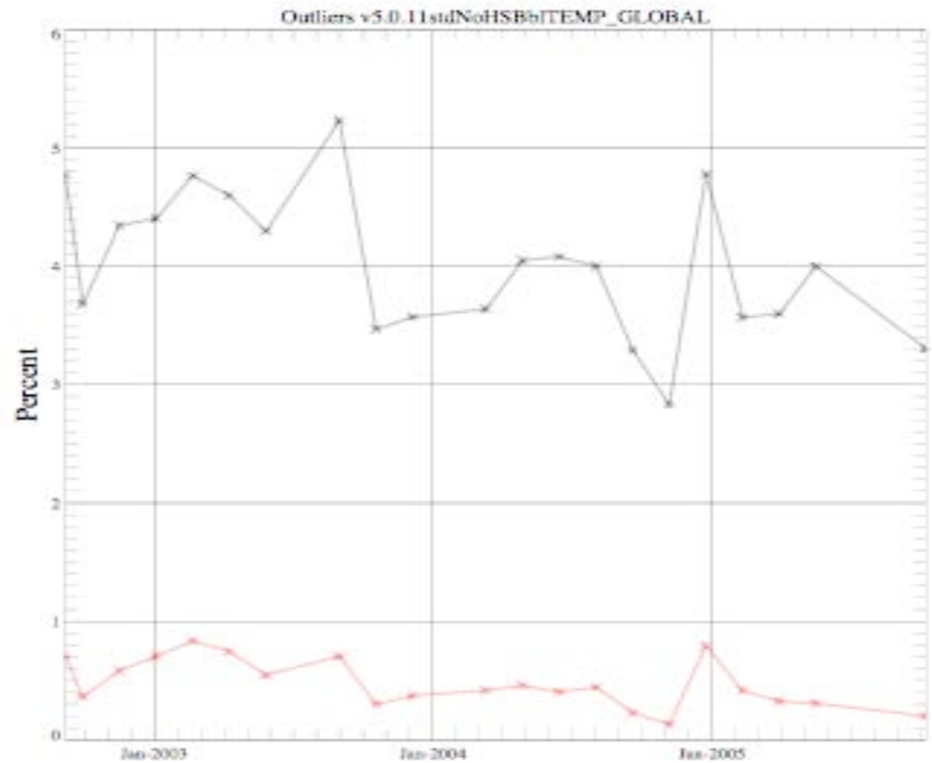


igure 14. Seasonal trends between AIRS retrieval bias 150 hPa to surface and CMDL MBL CO_2, 90°N–90°S. Average differences between RAOB and AIRS temperatures are indicated by solid circles, smoothed differences using a 2-month sliding boxcar average are indicated by the dashed line, and zonally weighted linear least squares fit for the CMDL MBL product are indicated by the solid line.

From Hearty - Trend in V5 Global Temperature

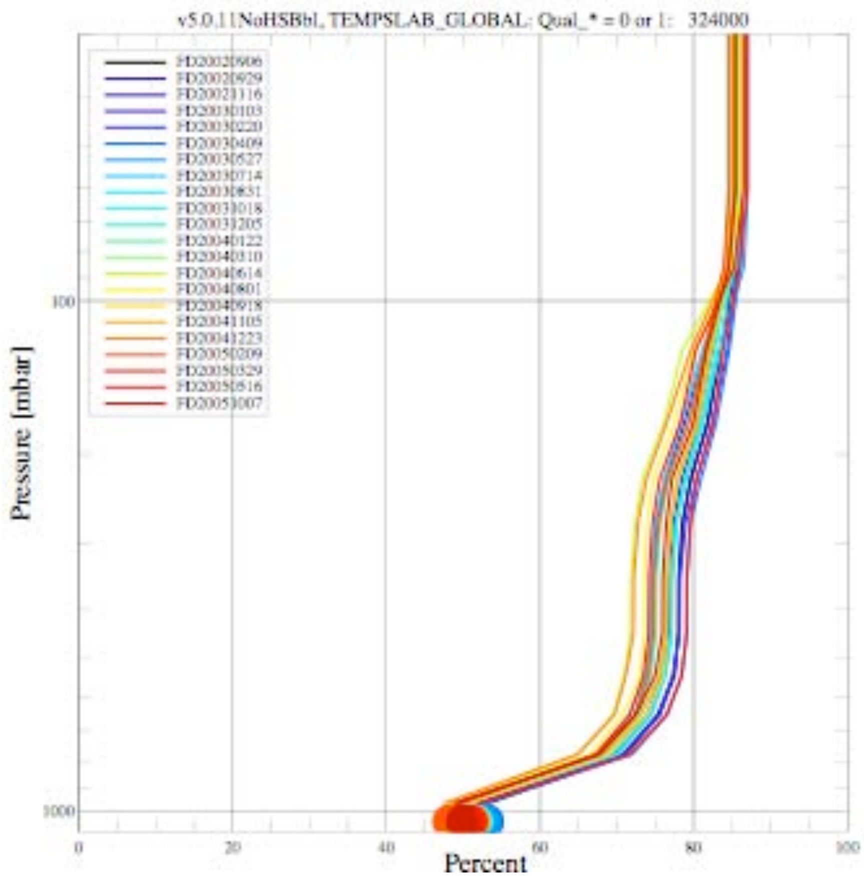
- **Upward trend in temperature bias vs. ECMWF**
- **Downward trend in outliers**

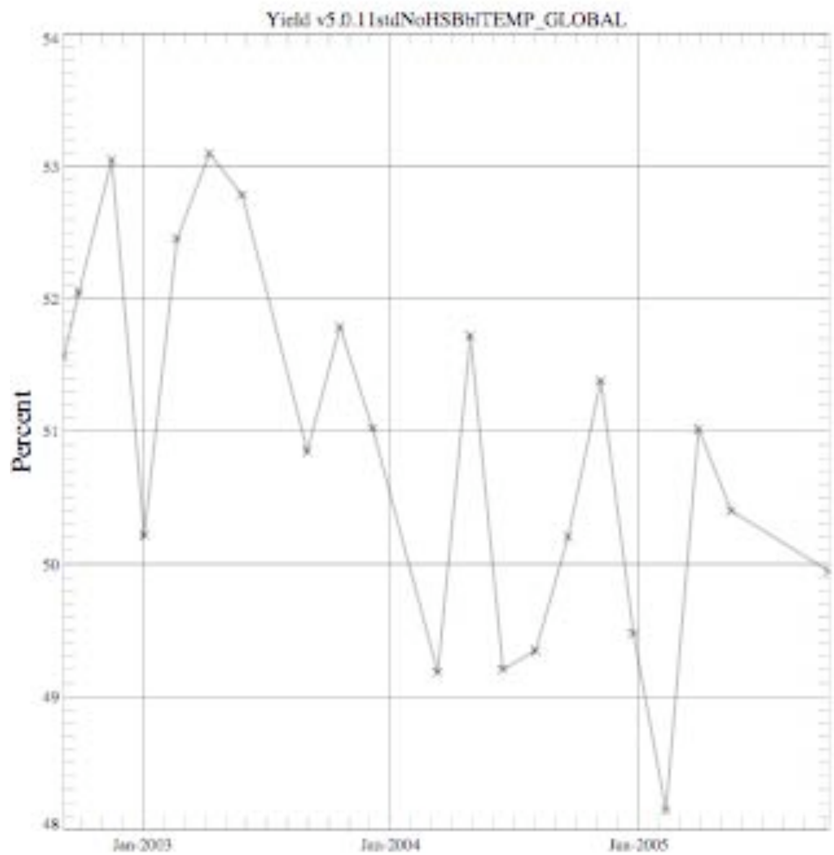




- Black line are mild outliers
- Red line are extreme outliers

From Hearty - Trend in V5 Global Temperature Yield





Much more in Hearty presentation in http://airs.jpl.nasa.gov/Science/ResearcherResources/MeetingArchives/TeamMeeting20070327/

Temporal Variation in Local Angle Adjustment

Background:

- AIRS v5 retrievals are performed over a 3x3 array of FOVs, assuming all differences among the 9 FOVs are due to clouds

- Because of the instrument scan pattern, these 9 FOVs are observed at 3 different angles through the atmosphere, introducing small differences in the spectra

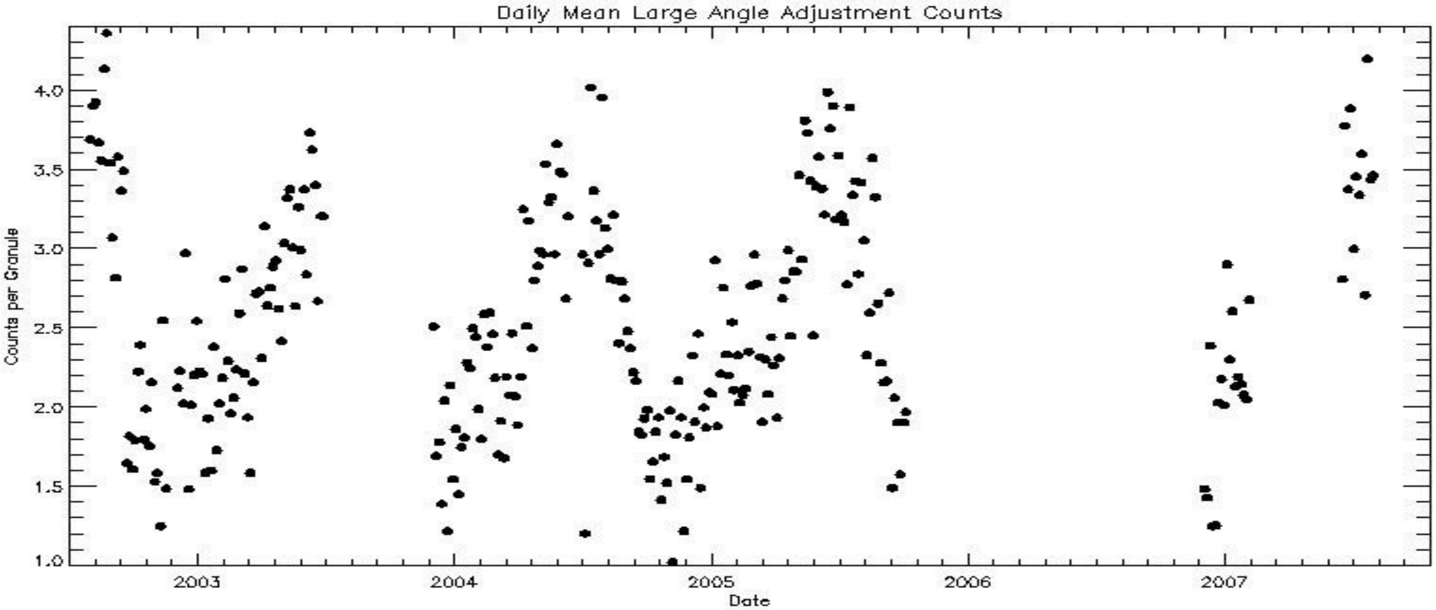
- Local angle adjustment makes small changes to the spectra from the outer 6 FOVs to emulate what would have been seen at the central angle

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

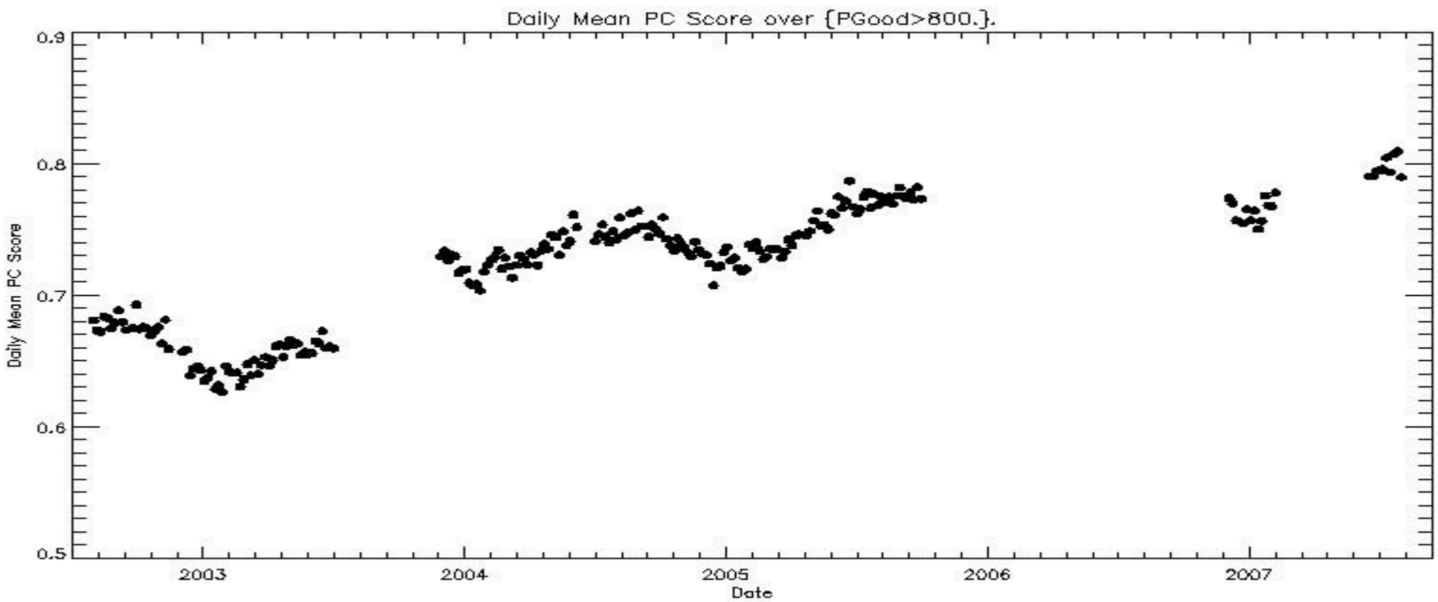
Temporal Variation in Local Angle Adjustment

- **Each 6-minute granule produces a count of number of FOVs with "big" angle adjustments (at least 5 channels adjusted by at least 20 * noise)**
- **The number of these cases shows a strong annual cycle**
- **But remember, LAA is a small adjustment (generally)**



Temporal Variation in PC Scores

- **Lower PC score means the input matches the training set better**
- **PC Scores are rising with time**
- **There is a clear seasonal cycle**
- **PC Scores are used in quality control -- higher PC scores mean more rejections.**



Daily Mean of PC Scores where Pgood > 800 mbar

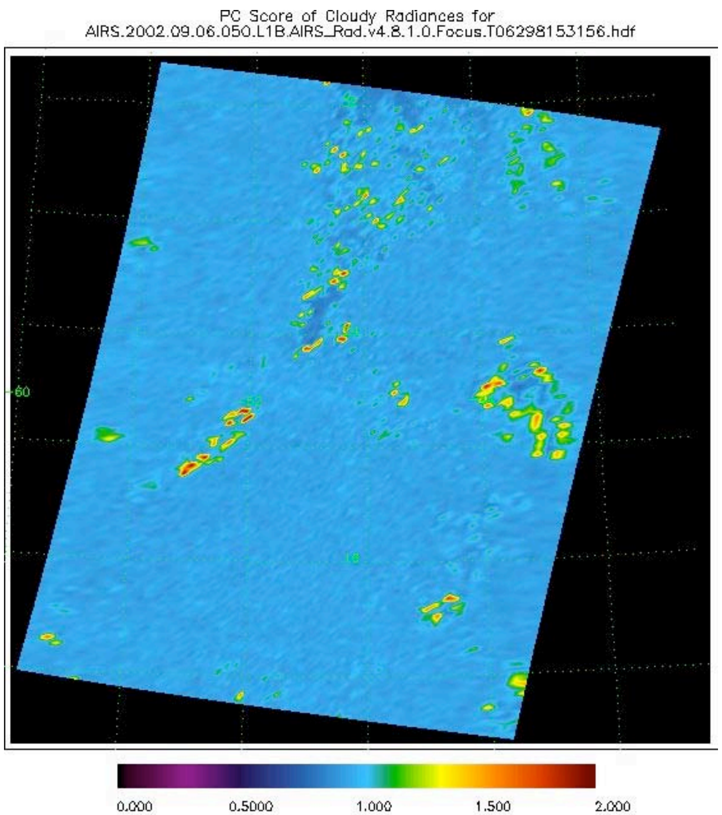
National Aeronautics and
Space Administration

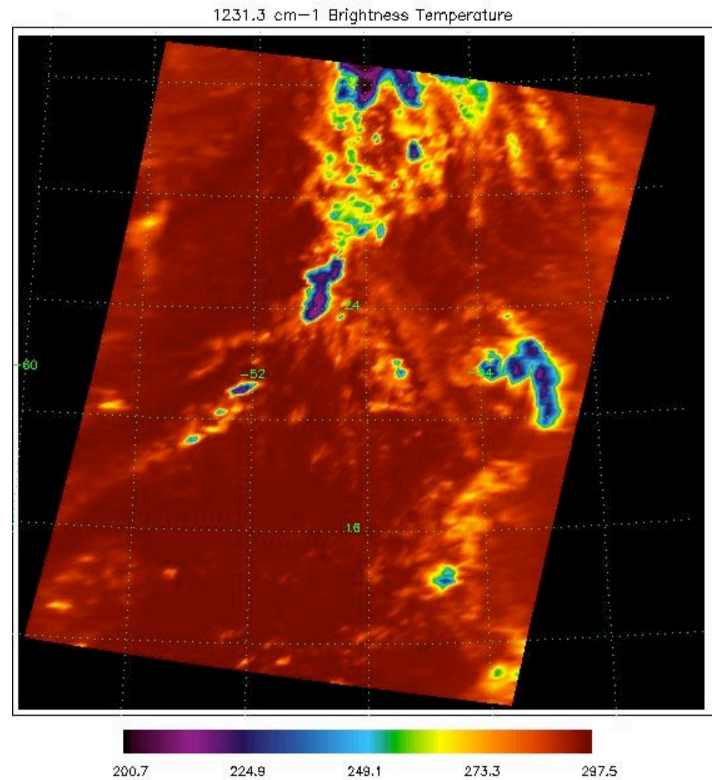
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Why Suspect Regressions?

- **Regressions occupy key points in the retrieval**
- **Regressions have a known dependency on training data -- they only know how to handle what they have seen before**
  - These regressions use a large number (~50%) of all 2378 AIRS channels. When any channel is unavailable, it must be filled somehow.
  - PC Scores are consistently elevated in regions of fires, dust, edges of clouds, sun glint, SO_2, etc.
  - Regressions are trained with a narrow range of background CO_2 will have trouble with later data with more CO_2

Difficult Cases for Regression -- Edges of Clouds

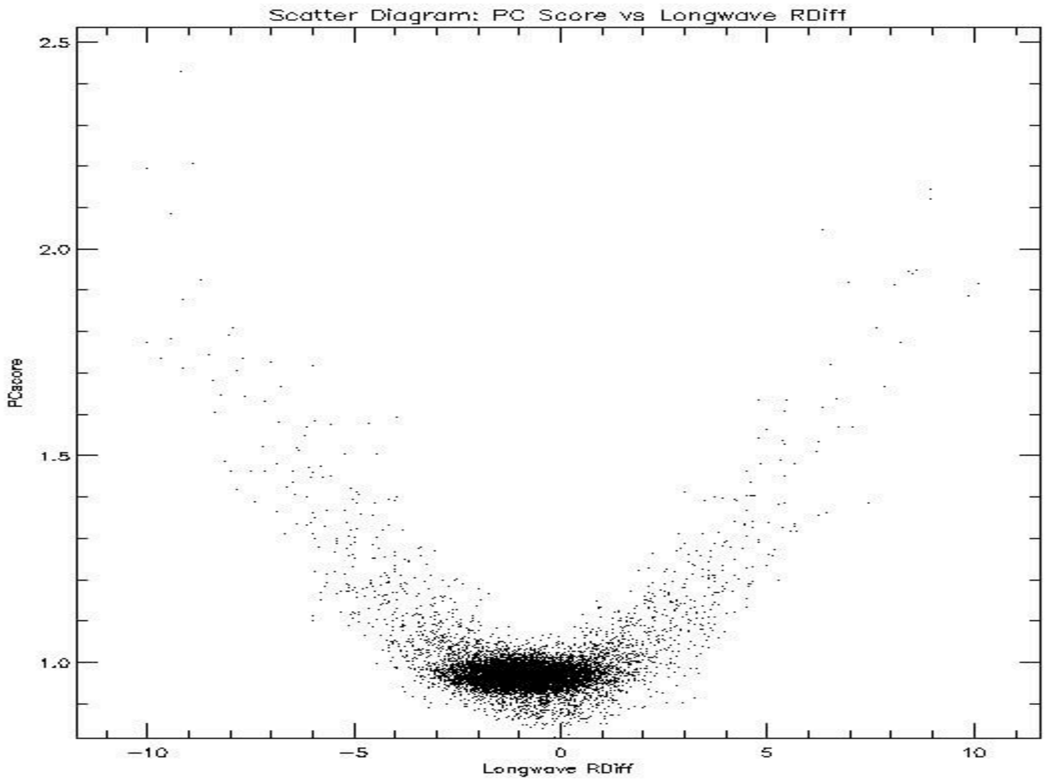




PC Score **Granule 50 of Sept 6, 2002** Tb 1231 cm^{-1}

High value of PC score is correlated with side of cloud, where C_{ij} tends to be high

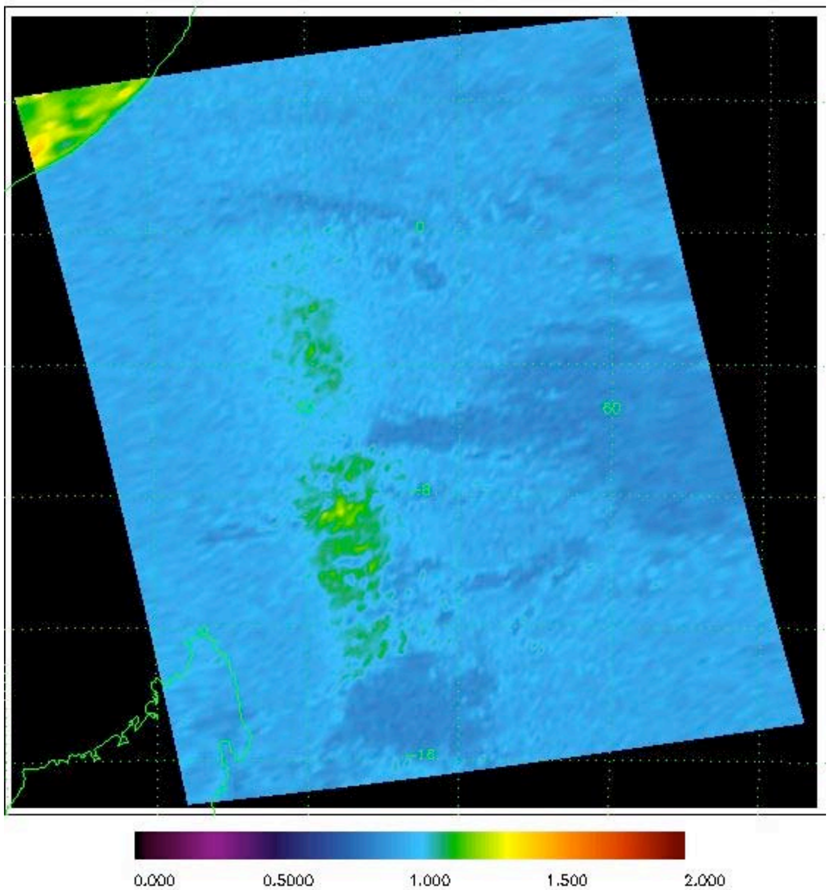
Difficult Cases for Regression -- Edges of Clouds (Cont)



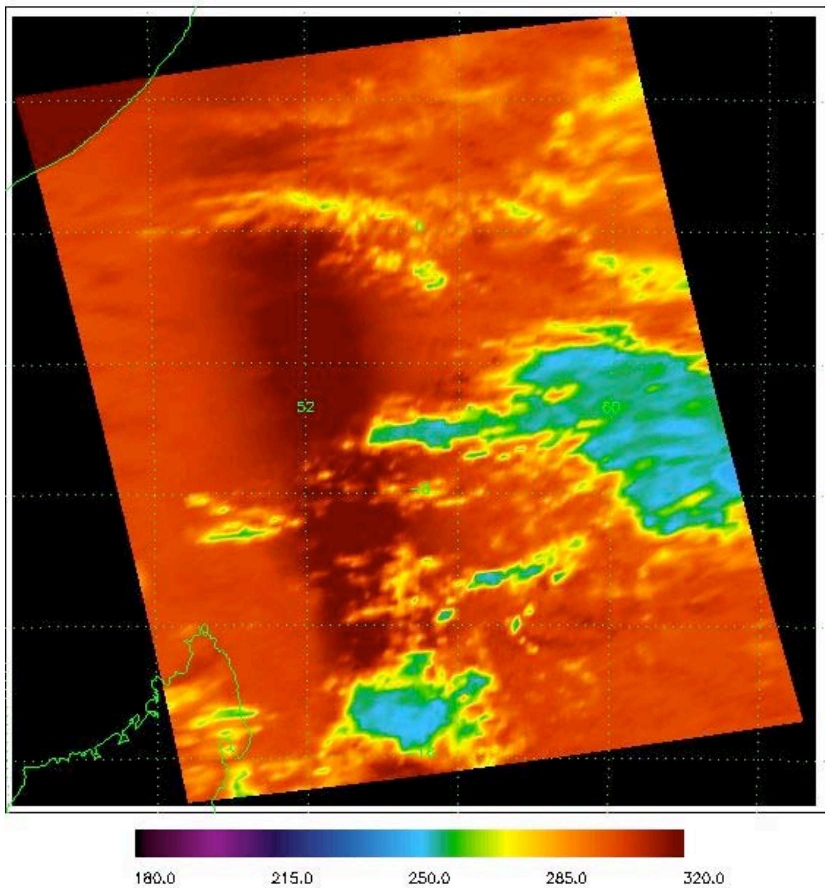
Scatter diagram of PC Score vs. Longwave Rdiff, a measure of C_{ij}

Difficult Cases for Regression -- Sun Glint

PC Score



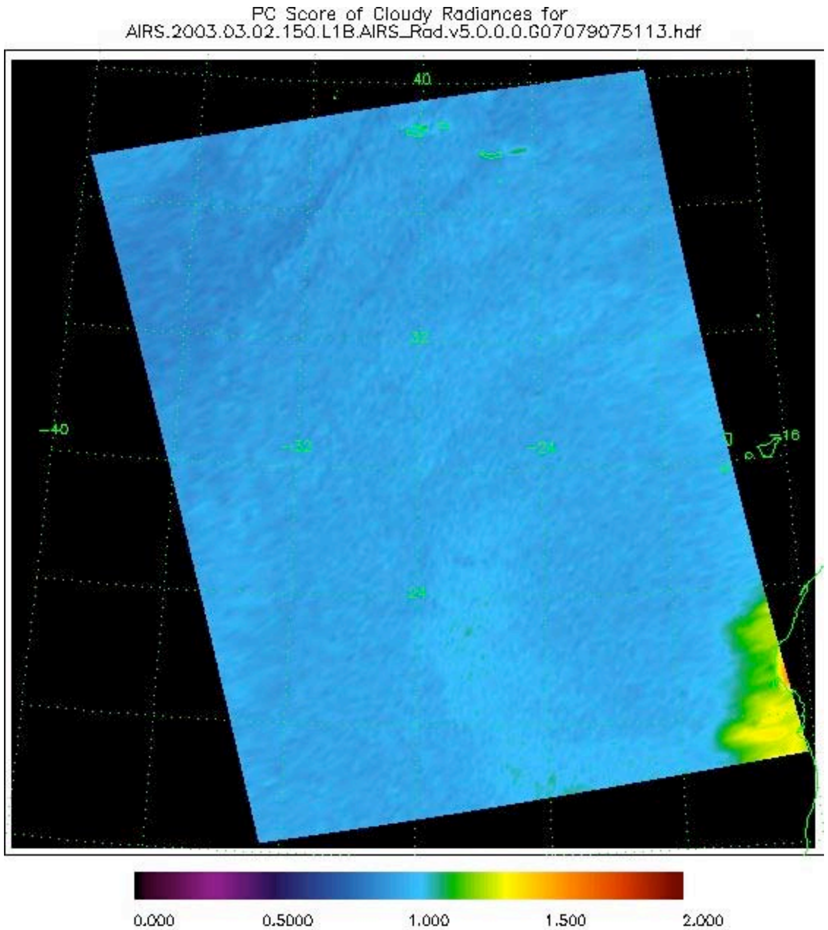
2616 cm^{-1} Br Temp



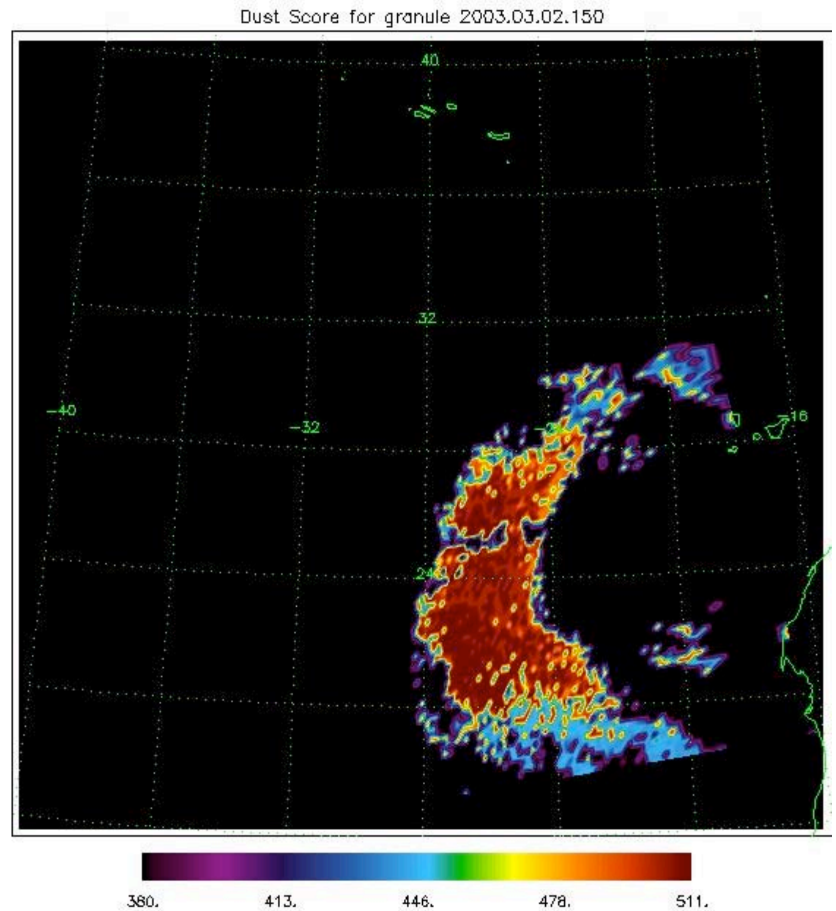
Granule 99 of March 2, 2003

Difficult Cases for Regression -- Dust

PC Score

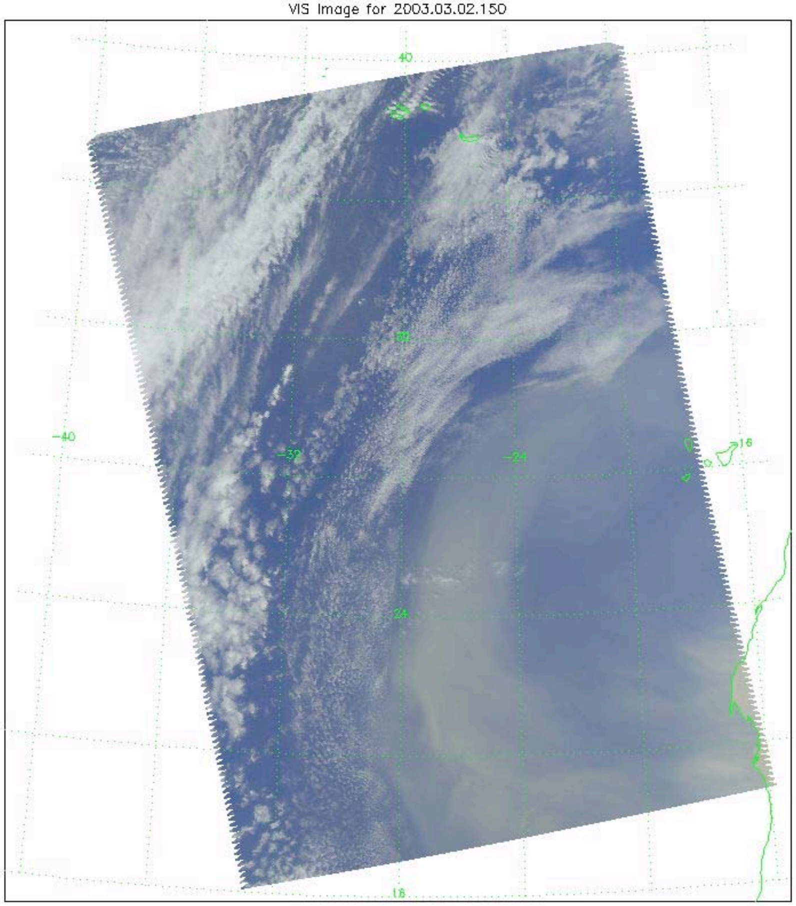


Dust Score



Granule 150 of March 2, 2003

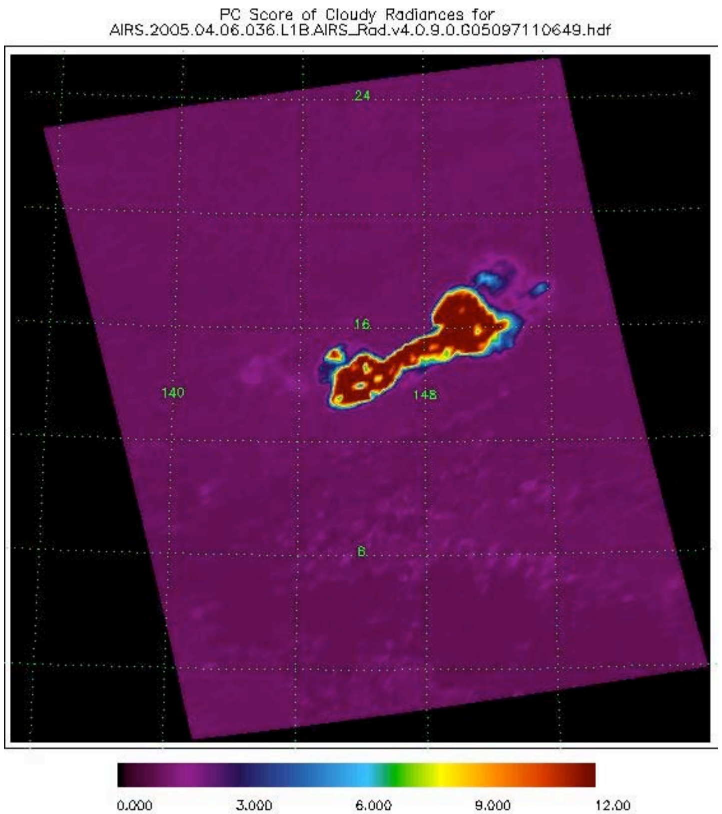
Difficult Cases for Regression -- Dust (cont)



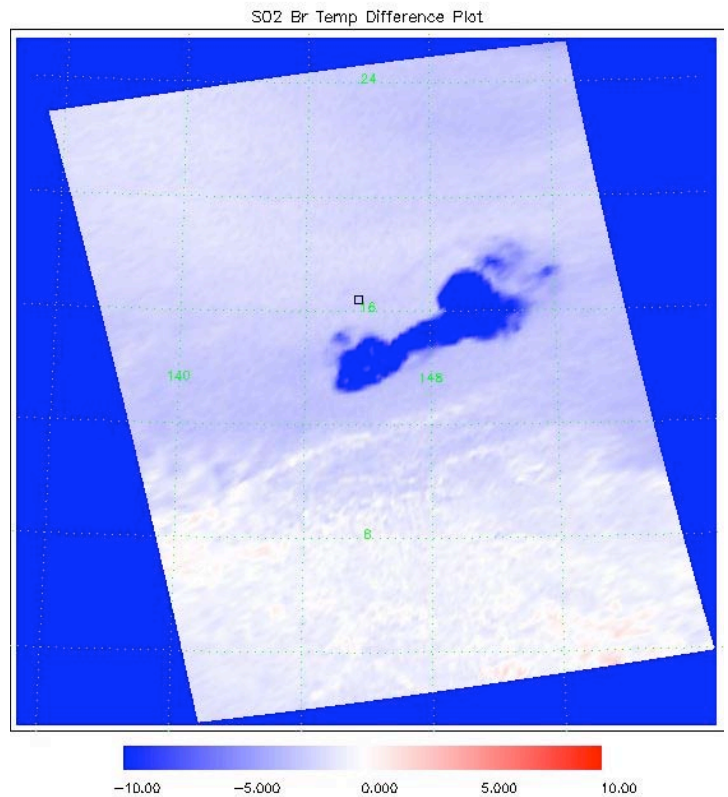
Dust Score Misses Some Dust

- **Dust plume near nadir is detected by dust score**
  - Only marginally high PC Score
- **Dust plume near the southeastern corner of granule is missed by dust score.**
  - Large PC Score

Difficult cases for Regression -- SO_2



PC Score

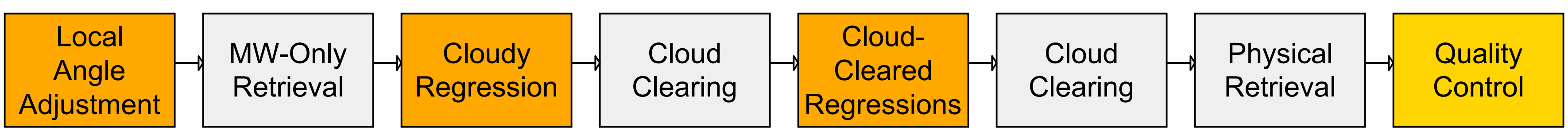


SO_2 Br Temp Diff

- **Volcanic plume from Anatahan**
- **Granule 36 of April 6, 2005**

Placement of Regressions

- **AIRS retrieval includes these key regression steps:**
  - Local angle adjustment
  - 1st guess cloudy regression
  - Cloud-Cleared profile regression
  - Cloud-Cleared surface property regression
- **Cloud Clearing plus physical retrieval as last retrieval step should attenuate the impact of upstream regressions**
- **Quality control mixes in regression results**
  - Uses PC scores
  - Uses differences between results of regressions and physical retrieval



Channel Filling

- Radiances of channels needed by regression are replaced with synthetic radiances when those channels are not considered useable.

- Overzealous standards have led to too many channels being filled. This will be reduced in version 6.

- The current channel filling algorithms are not optimal. They will be updated in v6.

- See details in backup material.

AIRS Channel Filling -- First 4+ Years

Year	Number of Channels Routinely Filled (out of 1680)
Late 2002-2003	1 - 7
2004	3 - 7
2005	2 - 14
2006-Early 2007	6 - 16

Spot check of 1st scan of granule #120 of selected focus days

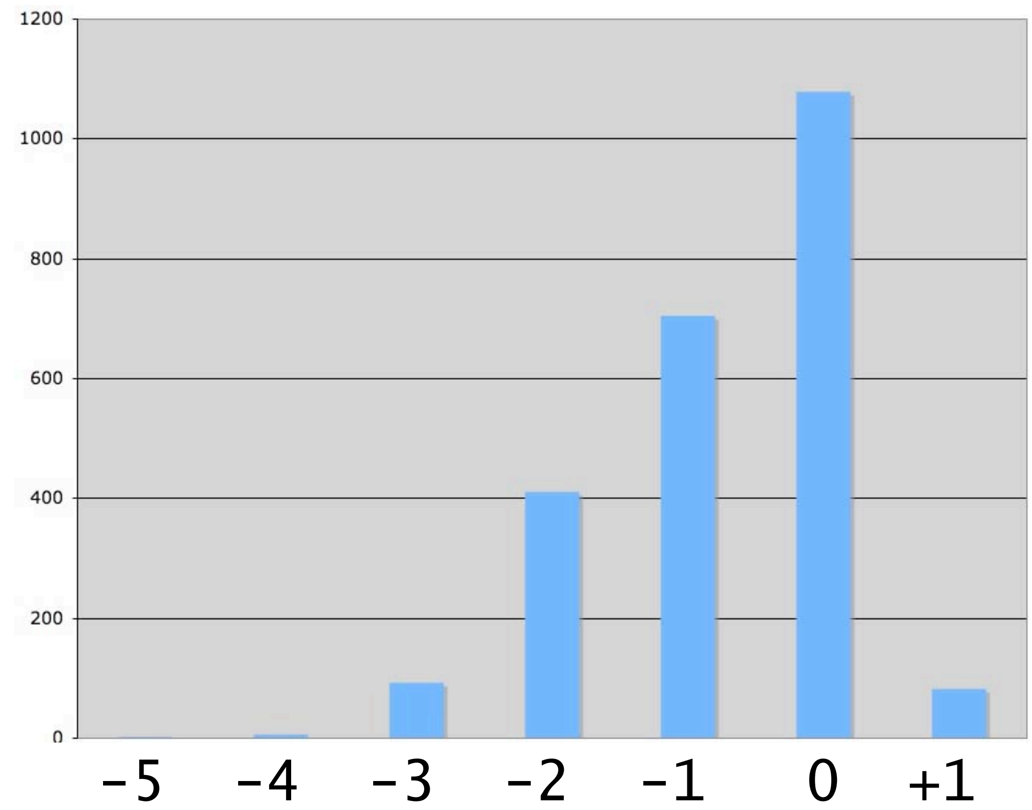
Tests of Channel Filling

- **These tests selectively block channels in Level-1B radiances and look at results of full retrieval**
  - Test 1
    - One granule is run 2378 times, with one channel flagged bad each run
  - Test 2
    - Data for 2002-09-06 (focus day 3) was run twice and results were compared:
    - 1st run is exactly released v5.0 product
    - 2nd run uses the v5.0 algorithm but the input is changed -- 15 channels which are not used on 2005-01-30 are flagged bad in the Level-1 input to retrieval

Results of Channel Filling Test 1

- **Histogram of change in yield of retrieval-type 0 (out of ~1000)**

Filling a single "average" channel causes yield to drop by ~0.1%



Results of Channel Filling Test 1

Channels with the largest effect on total cloudiness -- bias in *mean* cloudiness over an entire granule

Chan #	Freq cm^{-1}	CC 1	CC 2
2109	**2388.2**	**17 %**	**8 %**
2110	**2389.1**	**21 %**	**7 %**
1876	**2182.3**	**-7 %**	**2 %**
1871	**2186.9**	**-10 %**	**2 %**

- **The worst channels to lose are those near gaps in the regression set.**
  - Planned changes to the filling algorithms will fix this.
- **Physical retrieval reduces but does not eliminate the effect.**
- **Fortunately none of these channels have been lost.**

Results of Test 2

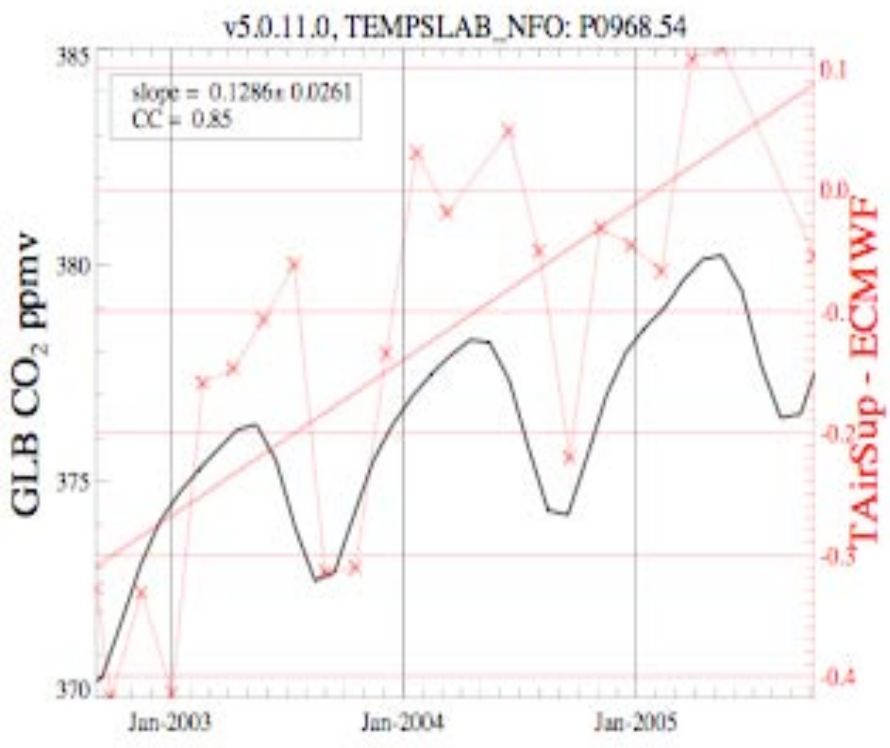
- **Differences caused by filling could be interpreted as climate trends**

Field	Change	Spurious Trend / 4 yrs
TSurfStd	+0.010 K	2.5 mK/yr warming
TSurfAir	-0.012 K	3 mK/yr cooling
TSurf1Ret	+0.026 K	6 mK/yr warming
O_3 yield	0.4% decrease	0.1 % /yr decrease
Initial_CC_score	0.0016	0.0004 /yr increase
nchan_big_ang_adj	-0.02	0.5%/yr decrease

- **Channel filling is not the main source of the spurious trends identified by Divakarla & Hearty (~100 mK/yr)**
- **But channel filling error is significant at the level of climate: ~10 mK/yr**
- **Effect on outliers not yet evaluated**

CO_2

- CO_2 changes with seasonal cycle and secular trend
- AIRS Level-2 trends resemble change in CO_2
- R-Branch CO_2 trend equivalent to ~50 mK/yr (per HHA)
- V5 regressions were trained on early mission data when CO_2 was lower
- V6 regressions will be trained to compensate



Hearty

Plans for v6

- **Limit use of regression**
- **Train regressions to handle the entire mission**
- **Improve channel filling**

V6 Plans -- Limit Use of Regressions

- **Angle Adjustment**
  - Evaluate simple training mean adjustment, segmented by angle and perhaps day/night, latitude, land/sea, solar zenith angle, etc. but not radiances
  - Joel Susskind is exploring 3x1 retrieval with no local angle adjustment

- **Cloudy regression, clear regression, surface regression**
  - Evaluate complete removal
    - MW first guess instead of cloudy regression
    - Mini-physical retrieval instead of clear regression
    - Surface emissivity guess from MODIS historical or climatology + MW for snow detection
  - Remove PC scores and differences of regression results from other retrieved states from error estimation and quality control

V6 Plans -- Train Regressions to Handle Entire Mission

- **Representative training sets to be selected by AIRS project and NOAA science team member**
  - Will be isolated from test data
- **Revisit channel selection to use only channels sensitive to target species**
- **May need multiple epochs to cover the entire mission**
  - This increases effort
  - Smooth transitions to avoid step functions at epoch boundaries
  - Must be careful of changes in models
- **Evaluate making regressions aware of CO_2**
  - Use time as a predictor
  - Use modeled CO_2 as a predictor

V6 Plans -- Improve Channel Filling

- **Will not fill as many channels**
  - Use values when noise level increases but is still under ~1 K
  - Eliminate lower limit of 150 K on cloud cleared radiances

- **Improve channel filling algorithms**
  - Evaluate alternate algorithms
    - First guess:
      - From all nearby channels, not just those used in regression
      - From channels selected for high correlation
      - From computed radiances based on the current state
    - Multiple passes through PC

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Backup Slides

California Institute of Technology

Jet Propulsion Laboratory

October 11, 2007

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration.

Cloudy and Clear Profile Regressions, Surface Regression

- **Cloudy regression was added in v5 as a partial replacement for MW-only retrieval**
- **It is used as a first guess of profile into the first iteration of cloud clearing**
- **Clear profile regression runs after first cloud clearing**
  - It provides a profile for use in the second iteration of cloud clearing.
  - Its fine vertical structure is preserved in physical retrieval retrieval
- **Surface regression runs right after clear profile regression.**
  - It provides an estimate of surface spectral emissivity used in second cloud clearing
  - Its fine structure is preserved through physical retrieval

Why Channels Are Filled

- **Noise levels of individual channels can change**
- **Some detectors have experienced significant long-term changes in noise levels**
    - See presentation by Denis Elliott
- **Other channels experience occasional changes in dark current ("pops") or transient high-noise events**
- **2-point calibration prevents any changes in bias -- only noise level changes**

Why Channels Are Filled

- **Training a regression implicitly makes it expect a given noise level -- it weights lower-noise channels more heavily**
- **Channels experiencing significantly higher noise levels than they had in the training set are not used as input to regressions**
- **But the regressions need input for all channels.**
- **Channel filling algorithms replace the radiance of a missing channel with a predicted radiance**
- **But note: the current screening of channels appears to be too strict, leading to too much channel filling. Version 6 will depend less on this method.**

Channel Filling

- **Different channel filling algorithms are used in different regressions.**
- **Local angle correction**
  - Initial radiance for filled channels set to training mean
  - PC scores calculated from radiances (including filled)
  - New radiances calculated from PC scores
- **Clear and cloudy profile regressions + surface regression**
  - Initial radiance for filled channels set to match mean of differences from training mean of radiances of 10 spectrally close* channels used in regression
  - PC scores calculated from radiances (including filled)
  - New radiances calculated from PC scores
- *** Channels selected may not be truly spectrally close because of gaps in the spectral coverage**

Pitfalls of Channel Filling

- **Filled values will not have correct noise characteristics**
- **Filled values will tend toward a training mean**
  - Output will tend to be correct in an average sense but extreme cases will be curtailed
- **Because the number of channels filled tends to increase with time, results will tend to systematically exclude extreme cases with time**

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Examination of PC Score

Sung-Yung Lee

California Institute of Technology
Jet Propulsion Laboratory

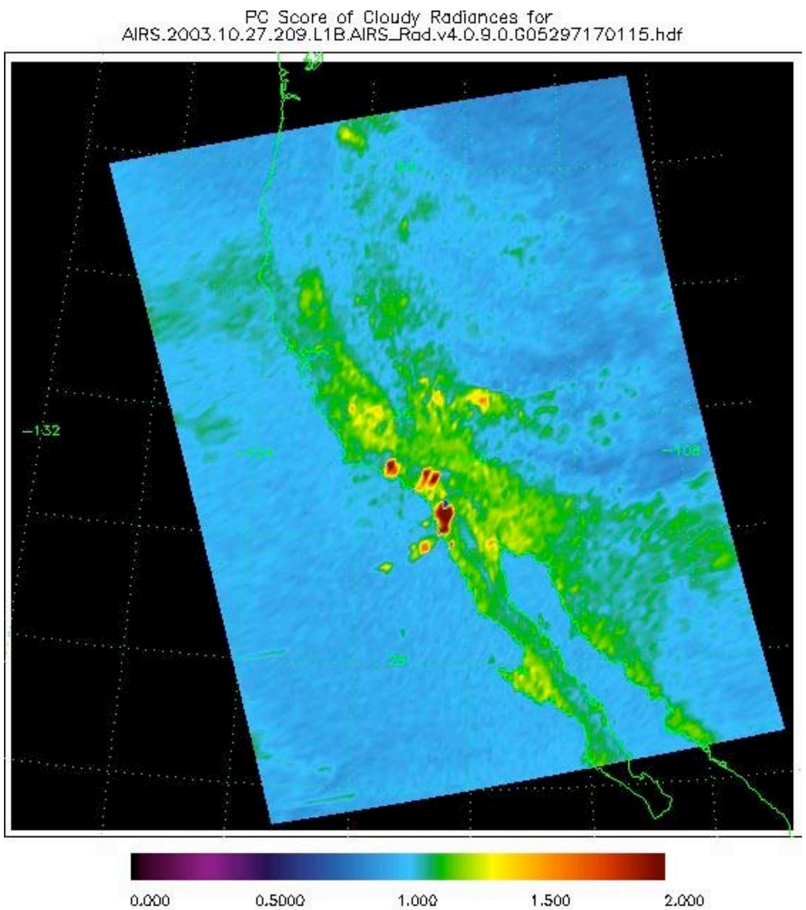
October 11, 2007

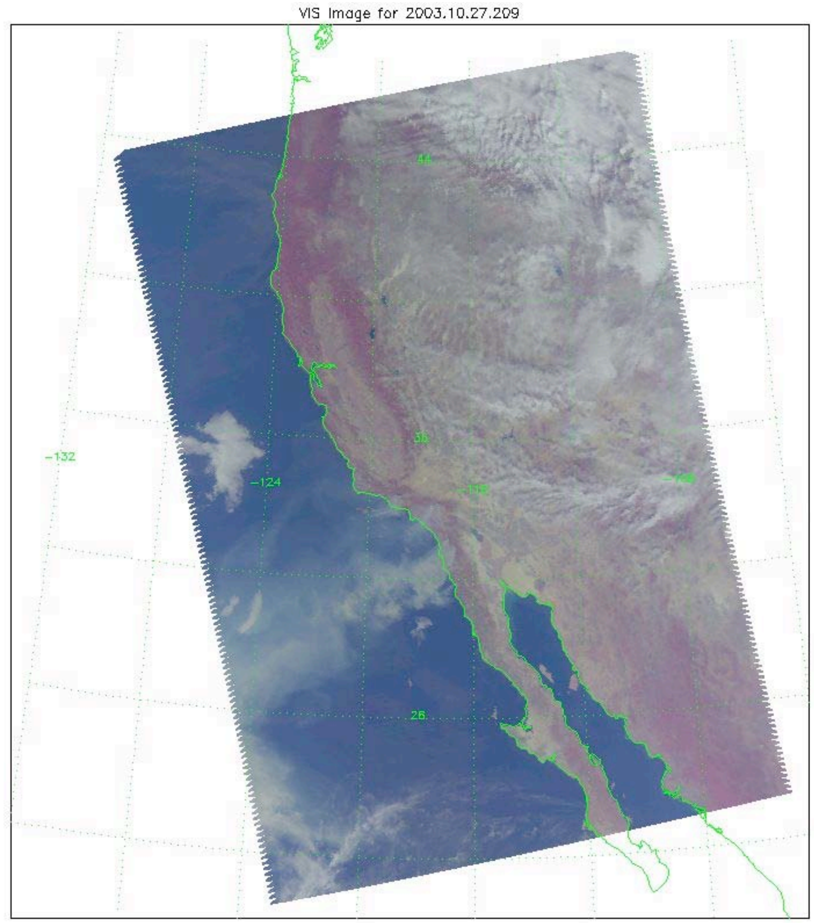
This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration.

Principal Component Score

- **M Goldberg (NOAA/NESDIS) developed algorithm to compress AIRS radiances as coefficients to principal components or eigenvectors.**
  - The principal components are computed from the radiances normalized by the channel noise, NeN.
  - Many of the channels used in the PC analysis became noisy over time
    - 25 channels as of mid-2007.
    - Early report claims the channel filling algorithm is reliable with fewer than 20 bad channels.
- **The Principal component scores is defined to be the residual error of the reconstructed radiances, in the unit of NeN.**
  - PC score of AIRS observed radiance is normally around 1.
  - PC scores are large when C_{ij} becomes an issue
  - PC scores are large over sun glint area and over brush fire
  - PC scores are large over “some” dust, but not all.
- **The initial regression algorithm of AIRS uses the principal components.**
  - It is applied to cloud cleared radiances.
  - The PC score is used as a measure of quality of cloud clearing.
  - Currently retrievals are rejected when PC score is larger than 4.

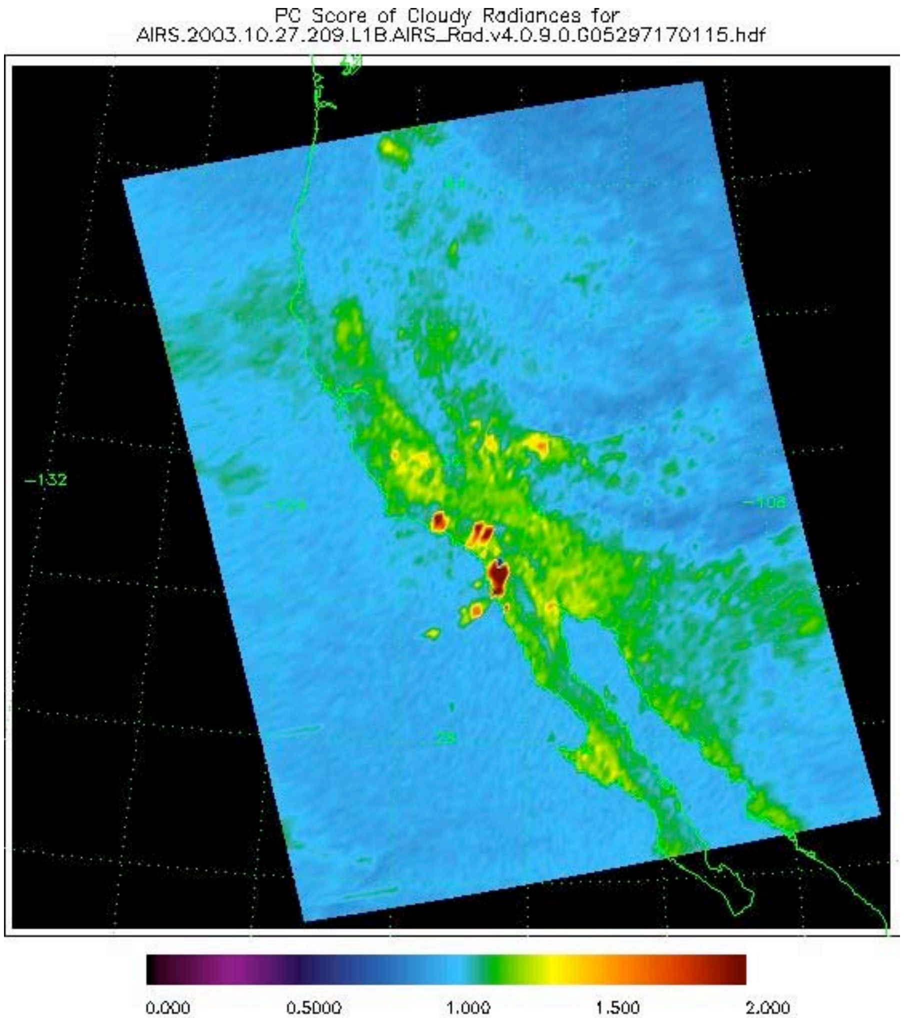
High PC Scores over Brush Fire





- **Granule 209 of Oct 27, 2003**

Granule 209 of Oct 27, 2003

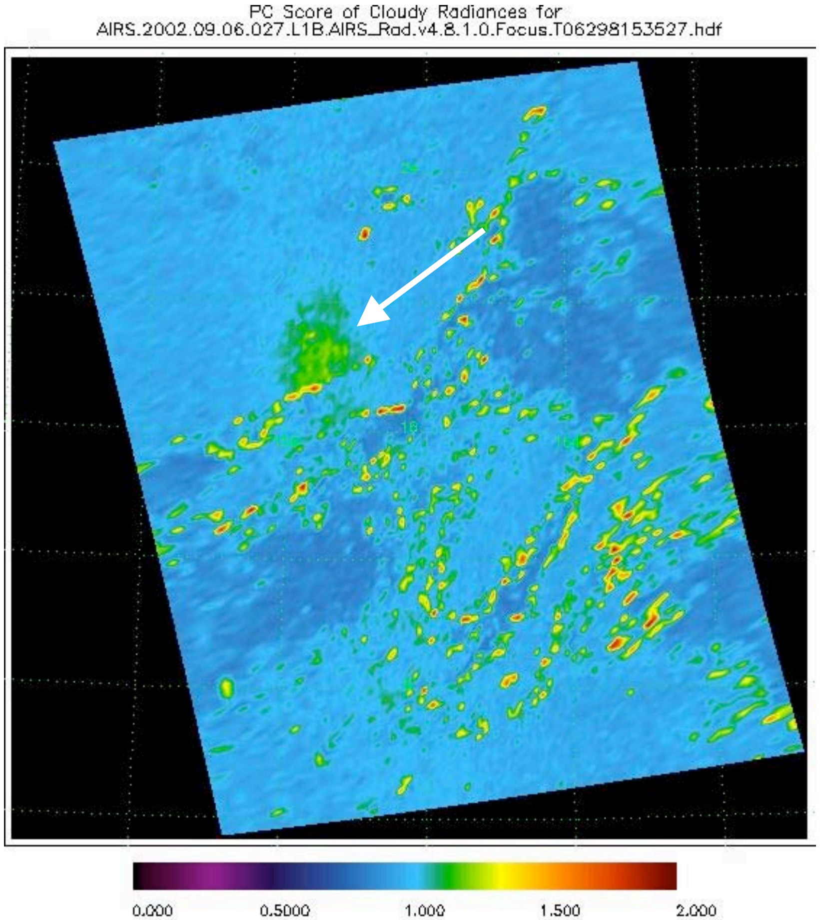


- **There are many brushfires burning in Simi Valley, Mt Baldy, Arrowhead, Cleveland Forest and San Diego.**
- **CO plume over ocean also has high values of PC score**

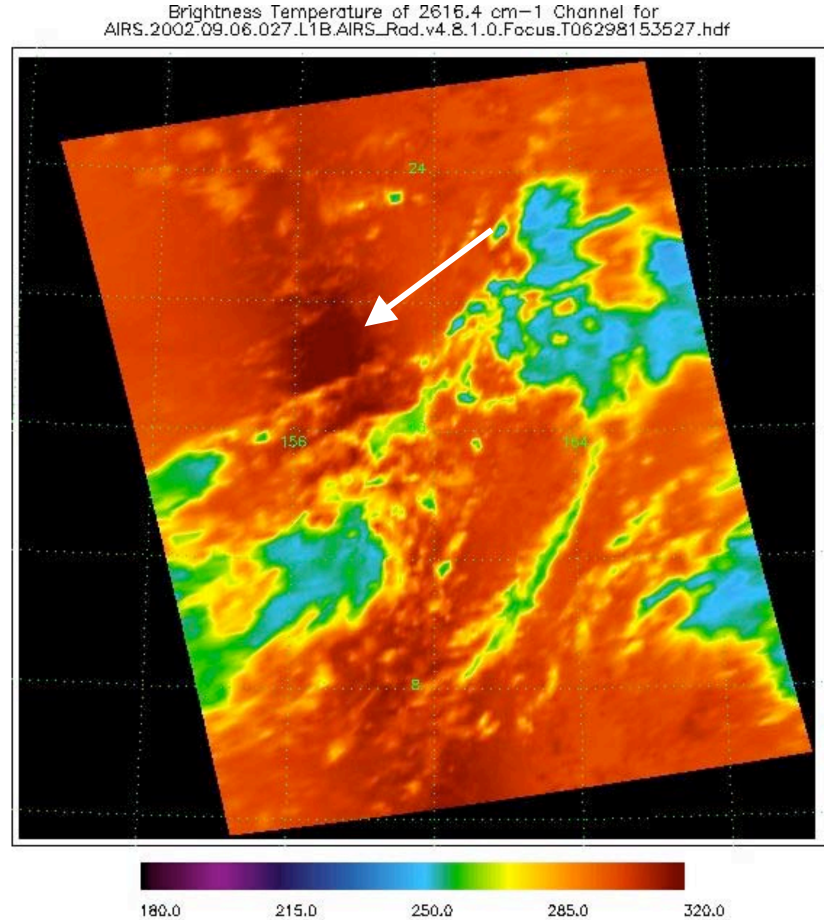
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Granule 27 of Sept 6, 2007



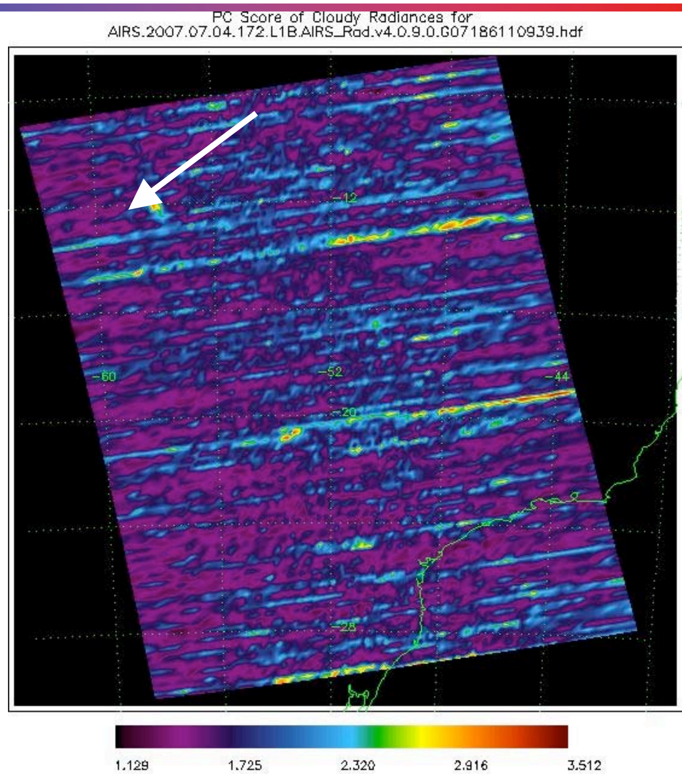
PC Score



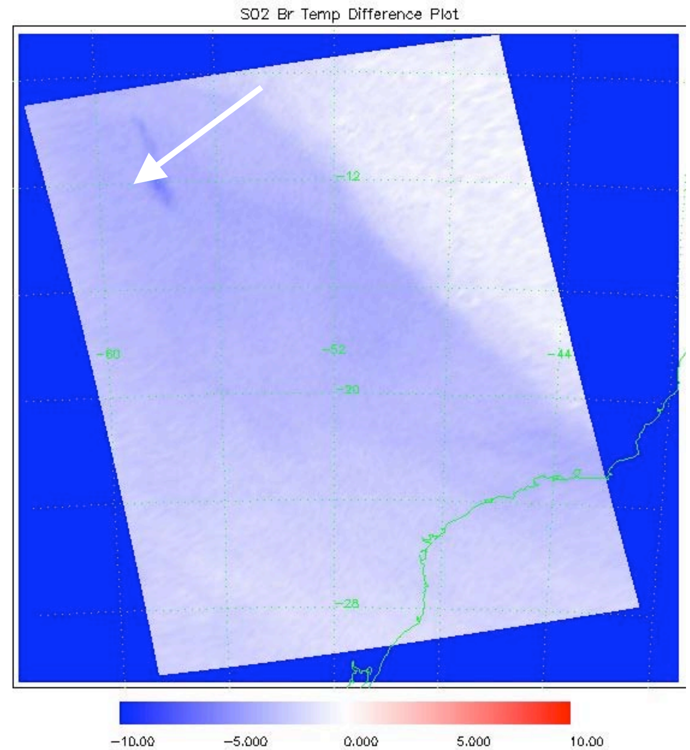
Tb 2616 cm^{-1}

- **Clear sun glint area has higher value of PC score**

Granule 172 of July 4, 2007



PC Score



SO_2 Br Temp Diff

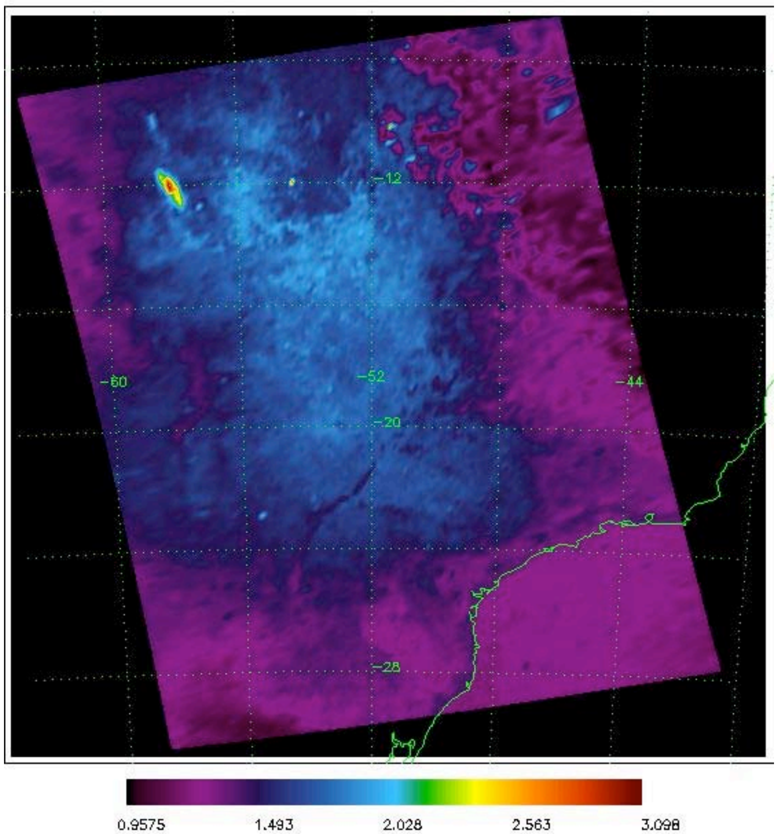
- **PC score (this version does not fill bad channels) degraded due to channel losses over time (29 bad channels in this granule)**
- **SO_2 plume (yet unknown source) is noticeable even in this figure**

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Granule 172 of July 4, 2007 (Continued)

PC Score of Cloudy Radiances for
AIRS.2007.07.04.172.L1B.AIRS_Rad.v4.0.9.0.G07186110939.hdf



- **Bad channels are filled using algorithm developed for local angle correction**
  - Fill bad channels with training mean
  - Reconstruct radiances using PC analysis
  - Fill bad channels with reconstructed radiance
  - Do another PC analysis to reconstruct
  - Compute PC score based only on good channels
- **The filled PC score clearly shows SO_2 plume**

Contingency -- Use Level-1C

- **Level-1C product will contain radiances resampled to a fixed grid.**
  - This eliminates a minor issue with regressions
- **Level-1C may also include filled values**
  - This would eliminate the need for channel filling in Level-2
  - Level-1C now would be responsible for avoiding the pitfalls of channel filling
- **Following this path for version 6 brings scheduling complications:**
  - Define Level-1C algorithm
  - Implement Level-1C
  - Run Level-1C on large dataset
  - Test Level-1C
  - If Level-1C acceptable, retrain all Level-2 regressions
  - If Level-1C is not acceptable, do something else in Level-2
  - Test Level-2